INTERNATIONAL JOINT CONFERENCE ON BIOMETRICS (IJCB 2022)

PDVN: A Patch-based Dual-view Network for Face Liveness Detection using Light Field Focal Stack

Yunlong Wang[1*], Mupei Li[1*], Zhengquan Luo[1,2], Zhenan Sun[1]
[1]Institute of Automation, Chinese Academy of Sciences (CASIA)
[2]University of Science and Technology of China (USTC)
yunlong.wang@cripac.ia.ac.cn, http://www.cripacsir.cn/

Abstract

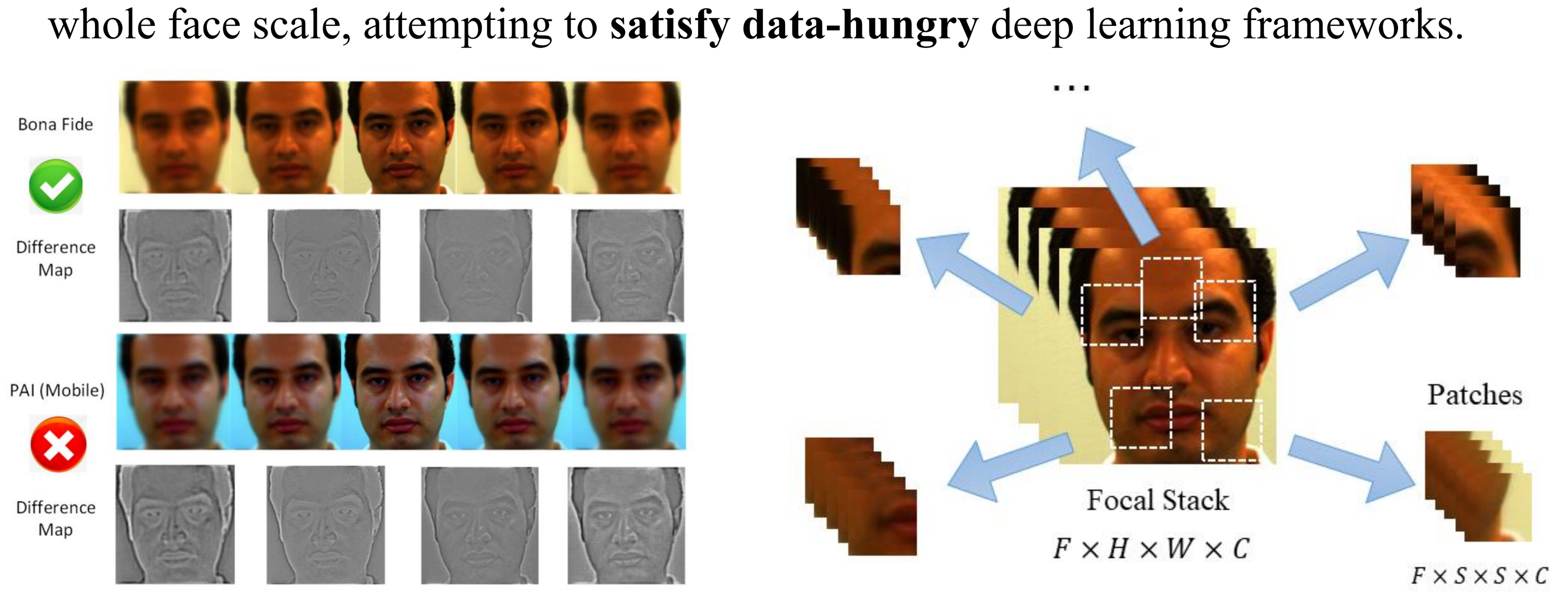
Light Field Focal Stack (LFFS) can be efficiently rendered from a light field (LF) image captured by plenoptic cameras. Differences in the 3D surface and texture of biometric samples are internally reflected in the defocus blur and local patterns between the rendered slices of LFFS. This unique property makes LFFS quite appropriate to differentiate presentation attack instruments (PAIs) from bona fide samples. A patch-based dual-view network (PDVN) is proposed in this paper to leverage the merits of LFFS for face presentation attack detection (PAD). First, original LFFS data are divided into various local patches along spatial dimensions, which distracts the model from learning the useless facial semantics and greatly relieve the problem of insufficient samples. The strategy of dual-view branches is innovatively proposed, wherein the original view and microscopic view can simultaneously contribute to liveness detection. Separable 3D convolution on the focal dimension is verified to be more effective than vanilla 3D convolution for extracting discriminative features from LFFS data. The voting mechanism on predictions of patch LFFS samples further strengthens the robustness of the proposed framework. PDVN is compared with other face PAD methods on IST LLFFSD dataset and achieves perfect performance, i.e., ACER drops to 0.

Problem Formulation

- Face liveness detection is an indispensable module of face recognition, however diverse PAIs, especially those simulating 3D geometric structure of human face.
- Light Field(LF) imaging contributes to face anti-spoofing, which enables to capture 4D spatial-angular information, implicitly recording 3D geometric structure and reflectance property of the object.
- Currently there is nil work exploring the merits of both LFFS(Light Field Focal Stack, rendered from raw LF) and CNN, which proved to be efficient.

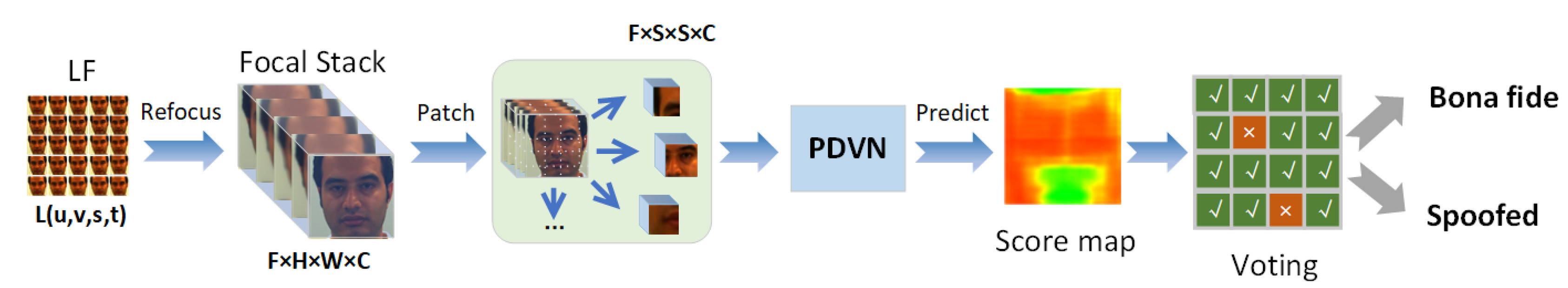
Patch LFFS Data Generation

- **Light Field Focal Stack.** Analyzing light intensity and directional information, the focus plane can be changed at any depth layers by the digital refocusing method proposed by Ng [44].
- **LFFS patches.** The spatial dimensions of patch LFFS data are cropped from a small portion of the whole face, aiming to exploit **local anti-spoofing clues** like color variation, texture and edge. Moreover, generating patch LFFS data exponentially increases the number of samples for training, which relieves insufficient samples of the whole face scale, attempting to **satisfy data-hungry** deep learning frameworks.



Left: Face liveness detection LFFS data. ***Right:*** Local LFFS patches.

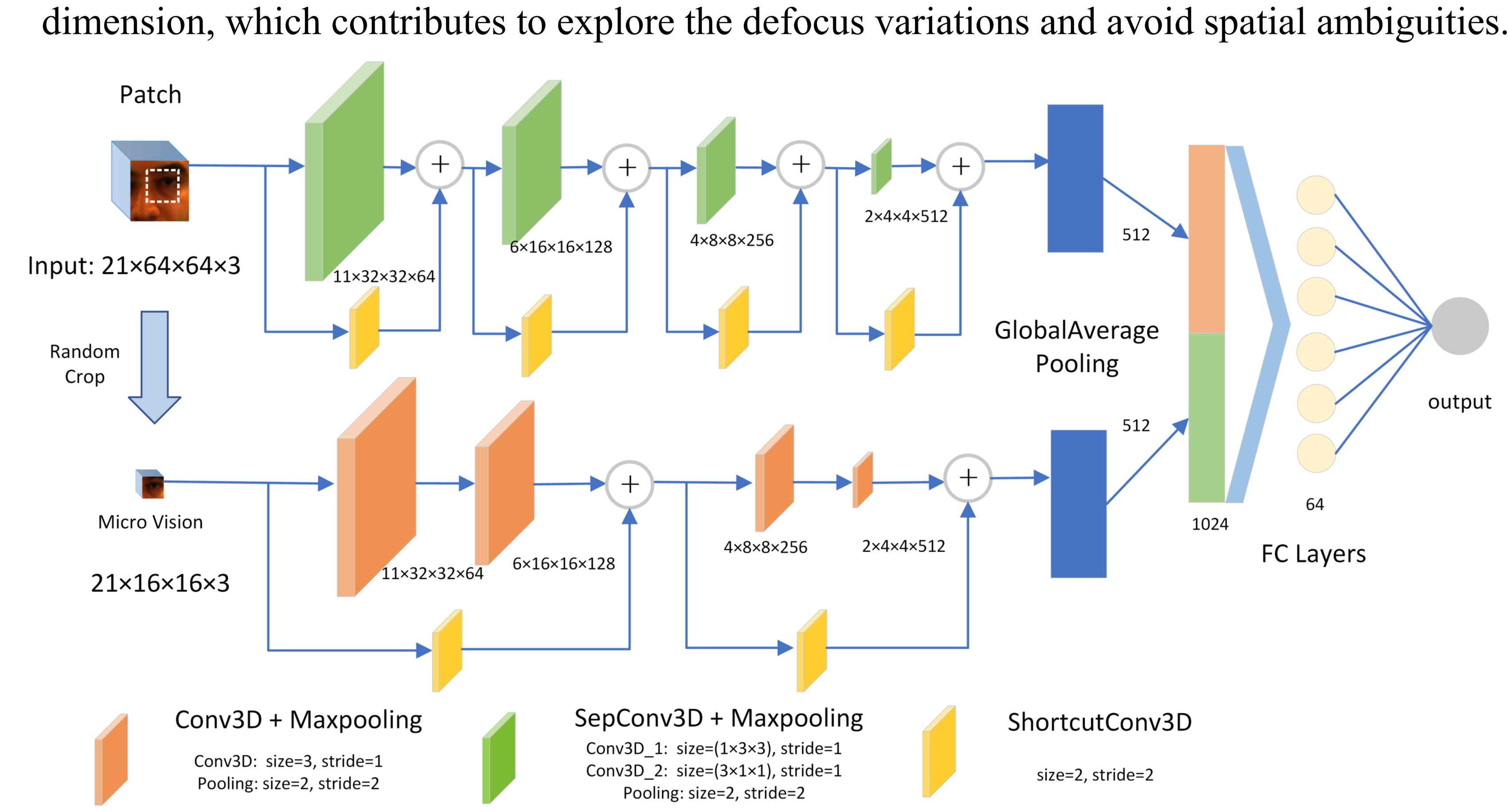
Framework



- **Local LFFS patches data.** To make fully utilization of original LFFS data, we divide them along spatial dimensions randomly, which drives the model to concentrate more on the intrinsic cues for liveness detection, rather than the layout of facial structure.
- **Specially-designed Network for local patches -- PDVN.**
- **Vote Mechanism.** The final detection result is accumulated by the vote of each patch, and the conclusion is made on the side with more votes, which further strengthens the robustness of proposed framework.

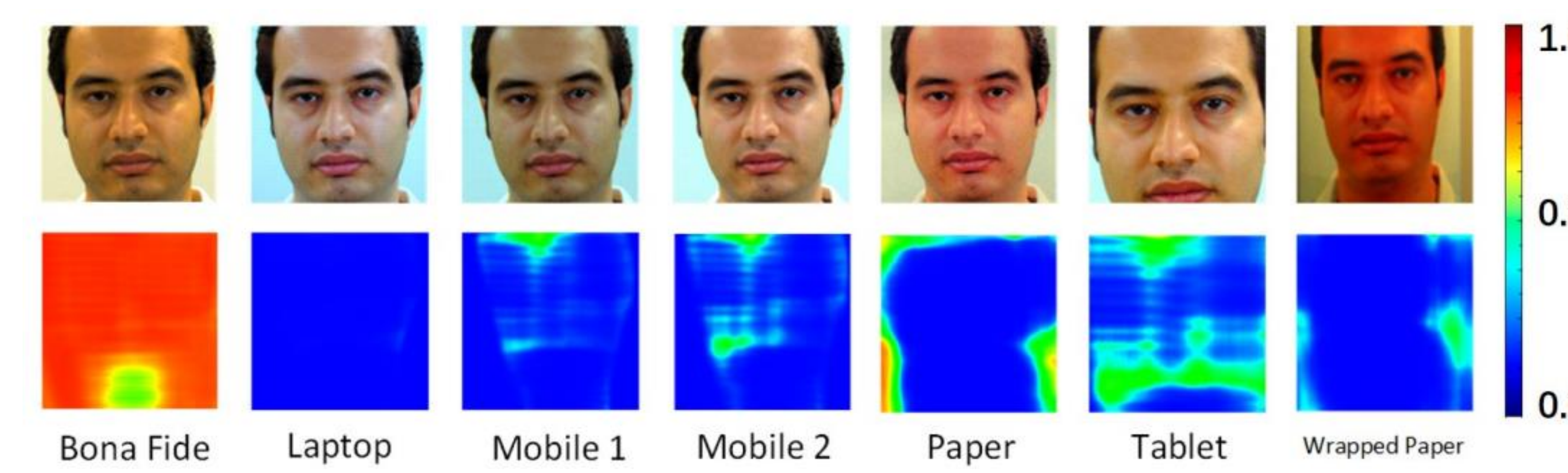
Proposed Network

- **Dual-view Branches.** We apply dual branches of different spatial resolution to gather more local region details, which highlights significance of combining original and microscopic view.
- **Separable 3D Convolution on Focal Dimension.** We split out convolution in focal dimension, which contributes to explore the defocus variations and avoid spatial ambiguities.



Above: The architecture of the proposed network (PDVN).

Demonstration of PAD



Above: All predictions of patch samples constitute the score map of the original LF sample. Red (close to 1) means real and blue (close to 0) means spoofing.

Experiment and Results

Compared methods	Lap	Tab	Mb1	Mb2	Pap	Wpa
2D image						
Määttä et al., 2011[1]	42.62	23.03	39.33	46.31	33.87	20.32
Määttä et al., 2012[4]	27.90	24.13	19.30	25.60	17.70	28.30
Wen et al., 2015[28]	12.06	13.01	9.23	15.83	14.82	15.27
Boulkenafet et al., 2016[6]	4.32	2.65	2.52	5.81	2.75	4.94
LF image						
Kim et al., 2014[20]	10.12	12.39	12.79	13.86	12.91	16.14
Raghavendra et al., 2015[21]	19.78	26.36	29.98	22.46	32.43	38.03
Ji et al., 2016[22]	11.00	10.77	8.12	18.50	7.27	22.05
Sepas-Moghaddam et al., 2018[39]	0.88	2.14	0.73	0.79	0.75	2.85
DL methods						
2D CNN	7.48	5.09	13.62	11.84	9.21	9.74
2D CNN fusion	3.80	4.47	4.35	3.80	3.80	4.95
CNN-LSTM	1.27	9.74	6.79	21.03	13.22	11.74
3D CNN	1.84	4.35	5.24	8.10	2.43	6.12
Ours	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**

Above: ACER of the proposed method and other compared PAD methods (%).

Conclusion

The proposed framework PDVN is validated to be a powerful DL based face PAD method using LFFS data, attributing to patch LFFS data generation, dual-view branches, separable 3D convolution on the focal dimension and voting mechanism on predictions of patch LFFS samples. Its superiority is experimentally verified on the widely adopted IST LLFFSD dataset. In future work, we will consider collecting databases and extending PDVN to other biometric traits, such as iris and fingerprint PAD.